DOCUMENT RESUME

ED 477 100 IR 021 828

AUTHOR	Suganuma, Akira; Nishigori, Shuichiro
TITLE	Automatic Camera Control System for a Distant Lecture with Videoing a Normal Classroom.
PUB DATE	2002-06-00
NOTE	7p.; In: ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (14th, Denver, Colorado, June 24-29, 2002); see IR 021 687.
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 3728, Norfolk, VA 23514. Tel: 757-623-7588; e-mail: info@aace.org; Web site: http://www.aace.org/DL/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Automation; *Computer Mediated Communication; Conventional Instruction; *Distance Education; Educational Technology; Higher Education; Instructional Innovation; *Lecture Method; Nontraditional Education; *Videotape Recordings

ABSTRACT

The growth of a communication network technology enables students to take part in a distant lecture. Although many lectures are conducted in universities by using Web contents, normal lectures using a blackboard are still held. The latter style lecture is good for a teacher's dynamic explanation. A way to modify it for a distant lecture is to capture the lecture by a video camera. When video lecture scenes are viewed, a camera-person usually controls a camera to take suitable shots; alternatively, the camera is static and captures the same location all the time. Both methods, however, have some problems. It is necessary to control a camera automatically. The authors are developing an ACE (Automatic Camera control system for Education) with computer vision techniques. This paper describes the system, the camera control strategy and an experiment of applying it to a real lecture. (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Automatic Camera Control System for a Distant Lecture with Videoing a Normal Classroom

Akira Suganuma Shuichiro Nishigori*

Department of Intelligent Systems,
Kyushu University, Kasuga, 816–8580, Japan
E-mail: suga@limu.is.kyushu-u.ac.jp

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract: The growth of a communication network technology enables us to take part in a distant lecture. Although many lectures are held by using Web contents in universities, normal lectures using a blackboard are still held. The later style lecture is good for teacher's dynamic explanation. All the way to modify it for a distant lecture is to capture by a video camera. When we video lecture scenes for the distant lecture, a camera-person usually controls a camera to take suitable shots; alternatively, the camera is static and captures the same location all the time. Both of them, however, have some defects. It is necessary to control a camera automatically. We are developing ACE (Automatic Camera control system for Education) with computer vision techniques. This paper describes our system, our camera control strategy and an experiment of applying it to a real lecture.

ED 477 100

1 Introduction

The growth of a communication network technology enables people to take part in distant lectures. There are various methods to held such a lecture. For example, it could be held by an web page-based method and a method of sending visual and audio of lecture scenes. One of the authors have been studying some supporting systems for a distant lecture. In the recent years, Suganuma et al. [Mine 2000, Suganuma 2000] have been developing a Computer Aided Cooperative Classroom Environment (CACCE) and an Automatic Exercise Generator based on the Intelligence of Students (AEGIS) for the web page-based lecture.

Indeed some lectures, for example the information technology or the programming, are frequently held by using visual facilities or computers in many universities but many lectures are held by the traditional style. It seems that such lectures will not disappear in the future although they will hold by combining the blackboard and a visual facility such as an OHP or Power Point software. We are, consequently, also developing another supporting system for the distant lecture with videoing the traditional lecture. This paper describes an Automatic Camera control system for Education (ACE).

We envisage that scenes of a lecture held in a normal classroom are recorded by a video camera and students in remote classroom take part in the lecture by watching the scenes projected on a screen. Figure 1 illustrates a form of the distant lecture by videoing the traditional classroom. A teacher teaches his students in an ordinary classroom. There is a blackboard in the room. He writes and explains something on it. Watching it and listening to his talk, students in the room take part in the lecture. Some cameras are setting in the room and take a lecture scene in order that the captured scene is sent to distant classrooms. On the other hand, students in the distant rooms take part in the lecture by watching a scene reflected on a screen.

When we video lecture scenes for the distant lecture, a camera-person usually controls a camera to take suitable shots, alternatively, the camera is static and captures the same location all the time. It is not easy, however, to employ a camera-person for every occasion, and the scenes captured by a steady camera hardly give us a feeling of the live lecture. It is necessary to control a camera automatically. ACE enables people to do it for taking suitable shots for a distant lecture. Receiving a scene from a camera, ACE analyses it and recognizes the complexion on the lecture. ACE judges what is important in the scene and controls the camera to focus on it. If the sent scenes are most suitable and effective, the educational effect on the students in the distant rooms is as good as that of the students in the room.

*Currently, Sony Corporation, Japan

IR021828

BEST COPY AVAILABLE

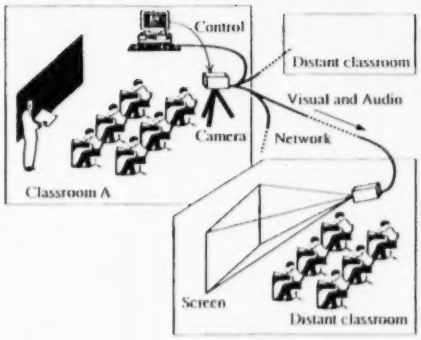


Figure 1: A form of the distant lecture by videoing the normal classroom

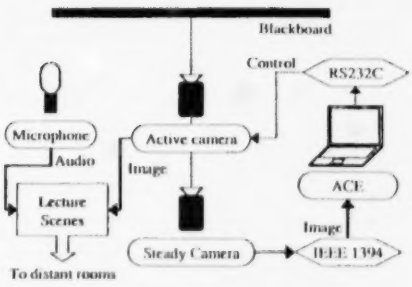


Figure 2: An Overview of ACE

In this paper, section 2 presents design of ACE and our strategy of camera control. Section 3 describes the algorithm to extract the latest object written by a teacher on a blackboard, and section 4 describes an experiment to apply ACE to a real lecture. Finally, concluding remarks are given in section 5.

2 Overview of ACE

2.1 Design

We have designed and implemented ACE, which is an application based on Computer Vision Technique. When we designed it, we assumed the following:

- A teacher teaches his students by using only a blackboard.
- Students aren't reflected in the scenes captured by the camera.
- A teacher isn't required to give the system a special cue.

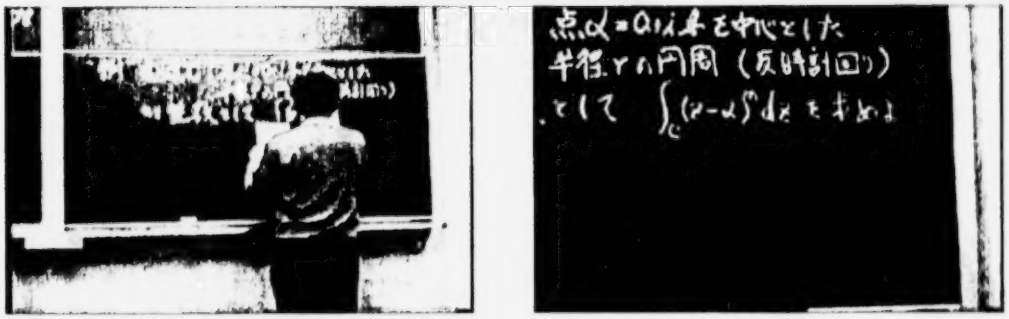
The first assumption means that the lecture captured by ACE is a traditional one. The teacher writes something on the blackboard, and explains them. Indeed he teaches his students using OHP and/or other visual facilities in the resent years, but many traditional ones are held in many schools.

The second assumption is made to decrease processing costs. If students are reflected in the scenes, ACE always has to distinguish a teacher and them. This processing is complex and take much time. It is easy to satisfy this assumption if we take a scene from the ceiling.

Finally, the third assumption is very important for a teacher. If a teacher gives ACE his special cue such as to press a button of a remote controller, ACE may control a camera more easily. ACE has only to keep waiting his cue in that situation. If the teacher, furthermore, put on a special cloth, on which some color markers are attached, it is easier to detect his position and/or action. The special cue and the special cloth, however, increase the load on the teacher. He may omit to give ACE his cue. He ought to concentrate his attention on his explaining. We decided, consequently, we didn't require him to give ACE his cue.

The overview of ACE is shown in Figure 2. ACE requires two cameras. One is a steady camera and the other is a active one. The steady camera captures a whole blackboard at a constant angle for image processing. The captured image is sent to ACE system running on a PC over an IEEE-1394 protocol. ACE analyzes the image and decide how to control the active camera according to a camera control strategy shown in section 2.2. The control signals are sent to the active camera over an RS-232C. The active camera, hereby, takes suitable shots. The visual and the audio are sent to the distant room. Students in the room watch and listen to them, and take part in the lecture. In our study, we are interested in how to video the lecture held in normal classroom. We are using a known method or product as a way sending the video via the network.

BEST COPY AVAILABLE



(a) An ordinary shot

(b) A key shot

Figure 3: Sample shots of a lecture scene captured by ACE

2.2 Camera control strategy

What does ACE capture? It is a very important thing for the system such as ACE. One solution is to take the scenes that students want to watch, but in this case many scenes are probably requested by many students at the same time. Although this solution needs the consensus of all students, it is very difficult to make it. We decide, therefore, that ACE captures the most important things from a point of view of a teacher. The most important thing from teacher's point of view is also difficult. We guessed the objects that teacher was explaining were the most important things for all students. When he explains something, he probably wants his students to watch it. He frequently explains the latest object that he have written on the blackboard. We decided consequently that ACE captured the latest object written on the blackboard.

Figure 4: A sample shot captured by a steady camera

When the lecturing scenes are videoed, both constantly changing shots and over-rendering shots are not suitable. A changeless shots are, if anything, more appropriate than those shots. It is important that students can easily read contents on the blackboard. The shots captured by ACE is shown in Figure 3. ACE usually takes a shot containing the latest object and a region near it in a discernible size. The blackboard often consist of four or six small boards like a picture in Figure 4. A teacher frequently writes relational objects within one board in this case. Now, ACE takes a shot by the small board such as figure 4-(a). On the other hand, ACE takes a shot zoomed in on the latest object after the teacher has written it on the blackboard such as figure 4-(b). After a-few-second zooming, ACE takes an ordinary shot again.

If we take the scene by a steady camera, the shot may be like a shot in Figure 4. In this case the camera must capture the whole blackboard, because the teacher writes something anywhere. The characters in this shot are too small for students to read. The shot of ACE is superior to that of the steady camera.

3 Extracting the latest object

3.1 Background subtraction

We use a background subtraction technique to detect objects on the blackboard. The background subtraction technique is a method to detach the foreground image from the background image. In the method, the background image is captured before opening the lecture. The image contains only the blackboard on which written no object.

BEST COPY AVAILABLE

It cannot contain a teacher. We can get some objects on the blackboard and in front of blackboard when we subtract the background from the image captured by the same camera during the lecture.

We adopted a background model[Haritaoglu 1998] in our system. This model is robust against a noise such as a flickered noise and so on. The platform is lightened by fluorescent lamps in a normal classroom. There are usually many noises such as flickered ones in a shot when a video camera captures objects lightened by them. ACE needs a robust method against noises for this reason.

In the original background model, the background image is modeled by representing each pixel by the following values; its maximum ($Max(p)$) and minimum ($Min(p)$) intensity values and the maximum intensity ($D(p)$) difference between two pieces of successive frames observed during capturing the background image. Pixel p is a background pixel if the intensity of pixel p satisfies the following: $Min(p) - I(p) \leq D(p)$ or $I(p) - Max(p) \leq D(p)$. The first inequality represents the lower-bound in order that the pixel belongs to the background, the other represents the upper-bound.

We specialized the method for the lecturing scene. The foreground objects segmented by the technique are something to write on the blackboard, something to erase on it, the teacher and so on. We need only the written object. Their pixels appear only above the upper-bound because the object written on the blackboard is brighter than the blackboard. Our method takes, therefore, no longer the pixel whose brightness is less than the lower-bound. ACE segments the object by using only the above second inequality.

The foreground objects are extracted by thresholding and noise clearing. The objects represent highlighted pixels in the background subtraction image.

3.2 Separating an object from the foreground image

The foreground image almost always includes a teacher. We would like to detect only the written object. If we mask teacher's region, we can get the region correctly. We have to detect, therefore, teacher's region. We assumed that all the moving object is a teacher. A method using a subtraction between a frame and a frame that captured after a short interval is usually used when we want to detect a moving object. ACE calculates the subtraction image and makes moving objects highlight. ACE makes a rectangle circumscribed highlight pixels the temporal teacher's region. After all pixels in teacher's rectangle in the foreground image are changed to dark ones, the remnant highlight pixels are the written objects if teacher's region is segmented correctly enough. ACE makes a rectangle circumscribed the highlight pixels, and deals with it after this processing.

3.3 Remake the background model

We have to distinguish the latest object and others. ACE keeps tracing the latest object written by a teacher from our camera control strategy described by section 2.2. Once the object has been detected as the written object, it doesn't need to be detected more than twice. After detecting the latest object, ACE re-calculates the values of the background model for each pixel in the region of the object. ACE always detects, therefore, only the latest one.

3.4 Timing of zooming in

We cannot control a camera even if we get the region of the latest object. We have to find the timing of zooming in. If ACE zooms in on the written object before a teacher has written, ACE must take a scene occluding the object behind the teacher body. After guessing whether the teacher finished writing, consequently, ACE zooms in on the object written by him at that moment.

The rectangle circumscribed the latest object usually change frame by frame. This main reason is the following:

- The rectangle increases or decreases because the teacher wrote something new or erased something.
- The masked region changes because the teacher moved to write something new. Then the rectangle increases or decreases.

Shortly, the rectangle changes when the teacher is writing something. On the other hand, he usually clears the object to make his students watch it after he has written. ACE take advantage of this feature to guess whether he finished. The rectangle does not change when he cleared the object. ACE counts the number of frames in which the rectangle does not change. If the number is over a threshold, then ACE judges the teacher finished writing, and control a camera to zoom in on the written object.

4 Applying ACE to a real lecture

4.1 Condition

We have developed ACE and done an experiment of applying it to a real lecture. We held two 25-minutes lectures on Mathematics for 85 undergraduates. A teacher taught them by only using a blackboard. One of the authors played the role of the teacher. Although he knows the detecting algorithm of ACE, he taught by usual style. Sort of thing, he behaved as disadvantage for ACE. We videoed the lecture scene by ACE and by a steady camera in order to compare their shots. These shots were recorded on video cassettes and played them in classrooms. We divided 85 students into two groups, each group discretely watched the video lectures projected on a screen. We evaluated ACE in terms of the following:

1. How correctly did the student copy contents from the screen to his note?
2. How good were the shots of ACE by comparison with that of the steady camera?

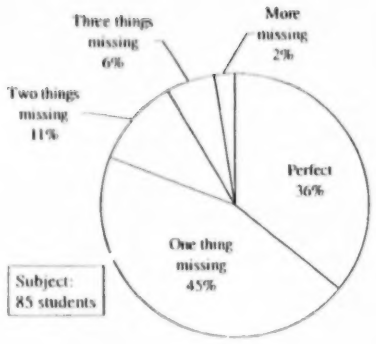


Figure 5: How many things did the student miss copying?

4.2 How correctly did the student copy contents?

We designed ACE to capture the latest object because a teacher often explains it. Is this strategy adequate enough to video a lecture scene? We evaluated the videoing strategy of ACE from a point of view that students could make a note. We made the students make a note for this evaluation. We told the students before the lecture that we would collect their notes after the video lecture and check that they ware made properly. We precomposed a master note and wrote the all contents of it on the blackboard all through the lecture.

After the lecture we collected their notes and compared them with the master note. We counted their missing in each note. The result is shown in Figure 5. 36% of all the students can perfectly copy the contents from the shots captured by ACE. 45% of the students make only one mistake. There are some small characters such as the index of a formula in this lecture scene. Many mistakes are caused by them. Almost all students make not more than three mistakes. Although there was a few students that hardly copied the contents from the screen to their note, this reason was that they sat dawn far from the screen when he took part in the video lecture. As the results of this estimation, almost all students can make a note correctly enough to learn it. The shots seem good enough.

4.3 How good were the shots of ACE?

We asked five questions to the students after each video lecture. They scored each scene captured by ACE and the steady camera, from 1 to 5. The average scores of each question are shown in Table 1. We omitted the third question for the scene captured by the steady camera because it always contains all objects on the blackboard. We used the t-test to compare these scores. Our null hypothesis is *"The scene captured by ACE is as good as the scene captured by the steady camera."*

The score of the scene captured by ACE is better then that of the scene by the steady camera, except the first question. Since ACE focus on the latest object, the scene does not always contain a teacher. Then the first evaluation of ACE is lower than that of the steady camera. On the other hand, a teacher is relatively small because the scene captured by the steady camera contains the whole blackboard. Teacher's action is hard to interpret in the scene. This is why the score of the steady camera is also low. Our null hypothesis, therefore, is not rejected with both 1% and 5% level of significance in this question. The point of the third question is, furthermore, what we abandoned at design of ACE. We were afraid that many students could not watch something to want because ACE zoomed in on the latest object. When they copy the contents, they often miss watching what and when they need, if timings of their watching do not synchronize with that of ACE's zooming in. Sure enough, the score is lowest of all scores of ACE.

BEST COPY AVAILABLE

Table 1: Questionnaire

Questions	Steady	ACE	Null hypothesis
1. Could you watch the teacher's action well?	3.27	3.19	Accept
2. Could you watch the objects on the blackboard well?	1.58	2.75	Reject
3. Could you watch the object you wanted?	—	2.49	—
4. Were you given a feeling of the live lecture?	2.71	2.86	Accept
5. Could you give the scene a overall score as a lecture one?	2.12	2.49	Reject

(1% level of significance)

Our null hypothesis is rejected with 1% level of significance in the second question and the fifth question. In both of them the scores of ACE are better than that of the steady camera. The evaluation of ACE is, therefore, superior to that of the steady camera. Neither of the scores, however, is high. In this experiment, the quality of the audio of our videos was not so good. Teacher's voice was, especially, hard to hear because we recorded it not by a wireless microphone at his breast but by a standard one on a desk. This is why the magnitudes of the evaluations are not large. If the quality had been better, they would have been larger. Because there were many feedbacks that described the low-quality audio. The shots captured by ACE is, consequently, good enough to video a lecture.

5 Conclusion

We have designed a camera control strategy for videoing a lecture and developed a prototype of ACE. We evaluated, moreover, it with applying it to a real lecture. As a consequence, we make sure that ACE is a useful tool for videoing a traditional lecture.

ACE takes a suitable shot if the teacher explains the object as soon as he writes on the board. It cannot take, however, a suitable shot when he explains something written in past. He usually teaches his students pointing the objects which he wants them to look at. Interpreting teacher's action and/or posture, ACE could capture more suitable scene. We will make ACE interpret it. We assume that a teacher teaches his students with a blackboard. But he sometimes also uses with an OHP. We will also make ACE be applied to such a situation.

Acknowledgments

This research was partly supported by the Grants-in-Aid for Scientific Research, Scientific Research Category C.

References

[Haritaoglu 1998] I. Haritaoglu, D. Harwood and L. S. Davis, "W^4: Who?When?Where?What? A Real Time System for Detecting and Tracking People," International Conference on Face and Gesture Recognition, pp.14-16, 1998.

[Mine 2000] T. Mine, A. Suganuma, and T. Shoudai, "The Design and Implementation of Automatic Exercise Generator with Tagged Documents based on the Intelligence of Students: AEGIS," Proc. of International Conference on Computers in Education, pp.651–658, 2000.

[Suganuma 1997a] A. Suganuma, S. Kuranari, N. Tsuruta, and R. Taniguchi, "An Automatic Camera System for Distant Lecturing System," Proc. of Conference on Image Processing and Its Applications, Vol.2, pp.566–570, 1997.

[Suganuma 1997b] A. Suganuma, S. Kuranari, N. Tsuruta, and R. Taniguchi, "Examination of an Automatic Camera Control System for Lecturing Scenes with CV Techniques," Proc. of Korea-Japan Joint Workshop on Computer Vision, pp.172–177, 1997.

[Suganuma 2000] A. Suganuma, R. Fujimoto, and Y. Tsutsumi, "An WWW-based Supporting System Realizing Cooperative Environment for Classroom Teaching," Proc. of World Conference on the WWW and Internet, pp.830–831, 2000.

BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

X	This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
	This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").